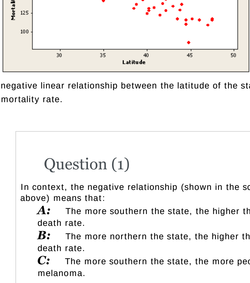
Case Câ†'C and Qâ†'Q Checkpoint

The first three questions refer to the following information:

The mortality rate from melanoma (skin cancer) during the 1950s was recorded for each of the 48 contiguous United States, plus Washington D.C. (as reported by Fisher and Van Belle (1993) and found on http://www.stat.psu.edu/~lsimon/stat501wc/sp05/data/).

The following is the scatterplot of the data:



The plot shows a negative linear relationship between the latitude of the state and melanoma mortality rate.

Question (1)

In context, the negative relationship (shown in the scatterplot above) means that:

A: The more southern the state, the higher the melanoma death rate.

B: The more northern the state, the higher the melanoma death rate.

C: The more southern the state, the more people died of melanoma.

Feedback

A : 10

✓ Good job! From the scatterplot we see that the lower the latitude the higher the mortality per 10 million. In other words, the more southern the state, the higher the death rate. Note that since the response variable is the mortality per 10 million, it measures the death rate rather than the number of deaths.

B : 0

✗ This is not quite right. Notice that there is a negative linear relationship between latitude and mortality per 10 million. In other words, as latitude increases, mortality per 10 million decreases. Consider the remaining options. (A) is the right answer.

C : 0

✗ This is not quite right. Notice that the variable mortality per 10 million is a rate, not an absolute count, of deaths from melanoma. Consider the remaining options. (A) is the right answer.

Question (2)

Based on the scatterplot and the value of the correlation coefficient, it would make sense to test the significance of this observed linear relationship between latitude and melanoma mortality rate. The appropriate hypotheses are:

A:

- H_o: melanoma mortality rate is related to latitude
- H_a: melanoma mortality rate is not related to latitude

B:

- H_o: melanoma mortality rate is not related to latitude
- H_a: melanoma mortality rate is related to latitude

C:

- H_o: melanoma mortality rate is linearly related to latitude
- H_a: melanoma mortality rate is not linearly related to latitude

D:

- H_o: melanoma mortality rate is not linearly related to latitude
- H_a: melanoma mortality rate is linearly related to latitude

Feedback

A : 0

✗ This is not quite right. The null hypothesis should assume no relationship. Consider the remaining options. (D) is the right answer.

B : 0

✗ This is not quite right. You are not testing for just any kind of relationship between the variables. Specifically, you are testing if there is a linear relationship between the variables. Consider the remaining options. (D) is the right answer.

C : 0

✗ This is not quite right. The null hypothesis should assume no relationship. Consider the remaining options. (D) is the right answer.

D : 10

✓ Good job! The hypotheses for testing the linear relationship between two quantitative variables is always: H_0: the explanatory and response variables are not linearly related. H_a: the explanatory and response variables are linearly related.

Question (3)

The following output is available:

Regression Analysis: Mort versus Lat

```
The regression equation is
Mort = 389 - 5.98 Lat


Predictor      Coef  SE Coef      T      P
Constant     389.19    23.81  16.34  0.000
Lat         -5.9776   0.5984  -9.99  0.000
```

From the output we learn that:

A: the data do not provide sufficient evidence to conclude that melanoma mortality rate is linearly related to latitude.

B: the data provide extremely strong evidence that melanoma mortality rate is linearly related to latitude.

C: the data provide moderately strong evidence that melanoma mortality rate is linearly related to latitude.

Feedback

A : 0

✗ This is not quite right. Notice that the p-value in the â€˜Latâ€™ row is extremely small (approximately 0). Consider the remaining options. (B) is the right answer.

B : 10

✓ Good job! The p-value is so small that Minitab reports that it is 0. This means that it would be essentially impossible to get data like those observed if melanoma mortality rate and latitude were not linearly related. This is extremely strong evidence that the two variables are linearly related.

C : 0

✗ This is not quite right. Notice that the p-value in the â€˜Latâ€™ row is extremely small (approximately 0). Consider the remaining options. (B) is the right answer.

The following four questions refer to the following information:

To determine if there is a relationship between grade performance and extracurricular participation, North Carolina state conducted a study of 112 students, recording the number of students in each of three extra-curricular categories, and each of two grade categories. Here are the results:

	Good grades	**Poor** grades
Low extra-curricular participation	11	2
Moderate extra-curricular participation	68	23
High extra-curricular participation	3	5

(*Source:* Felder, et. al., "A Study of Student Performance in an Introductory Chemical Engineering Course," *1992 ASEE Annual Conference Proceedings, pp. 1516–1519.)*

Question (4)

What is the expected count of students with high extra-curricular participation and good grades?

A: 3

B: 5.86

C: 8

D: 112

E: None of the above.

Feedback

A : 0

✗ That is not quite right. We find the expected number of students with high extra-curricular participation and good grades by assuming independence between extra-curricular activity and grades. Then multiply the probability of having high extra-curricular participation and good grades with the total

✗ number of people in the table. Consider the remaining options. (B) is the right answer.

B : 10

✓ Good job! The expected count is: (row total * column total) / table total = 8 * 82 / 112 = 5.86.

C : 0

✗ That is not quite right. We find the expected number of students with high extra-curricular participation and good grades by assuming independence between extra-curricular activity and grades. Then multiply the probability of having high extra-curricular participation and good grades with the total number of people in the table. Consider the remaining options. (B) is the right answer.

D : 0

✗ That is not quite right. We find the expected number of students with high extra-curricular participation and good grades by assuming independence between extra-curricular activity and grades. Then multiply the probability of having high extra-curricular participation and good grades with the total number of people in the table. Consider the remaining options. (B) is the right answer.

E : 0

✗ That is not quite right. It may be helpful to go back and review this section before answering this question. (B) is the right answer.

Question (5)

The count that you found in the previous question is the number of students with high extra-curricular participation and good grades that you would expect to see assuming that:

A: the conditions that allow us to safely use the chi-squared procedure are met.

B: extra-curricular participation and grade performance are related.

C: extra-curricular participation and grade performance are independent.

D: the null hypothesis of the chi-squared test for independence in this case is true.

E: both (C) and (D) are correct.

Feedback

A : 0

✗ That is not quite right. Remember that the count found above relies on the fact that the categories are independent. Consider the remaining options. (E) is the right answer.

B : 0

✗ That is not quite right. Remember that the count found above actually relies on the opposite assumption, that the categories are independent. Consider the remaining options. (E) is the right answer.

C : 0

✗ That is not quite right. Although this is a true statement, there is a better option. Consider the remaining options. (E) is the right answer.

D : 0

✗ That is not quite right. Although this is a true statement, there is a better option. Consider the remaining options. (E) is the right answer.

E : 10

✓ Good job! In general, the expected counts are the counts that we would expect to see if the null hypothesis of the chi-square test (which claims that the two categorical variables are not related, or independent) were true. In this case, then, the expected counts are calculated assuming that extra-curricular activity and grade performance are not related.

Question (6)

The following is the (edited) output of the chi-squared test in this case:

Chi-Square Test: Good, Poor

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

              Good    Poor  Total
Low             11       2     13
             9.52    3.48
             0.231   0.631

Moderate        68      23     91
            66.63   24.38
            0.028   0.079

High             3       5      8
                     2.14
            1.394   3.810

Total           82      30    112

Chi-Sq = 6.171, DF = 2, P-Value = 0.046
```

Based on the output (and using the traditional significance level of 5%) we can determine that:

A: the data provide sufficient evidence to conclude that extra-curricular participation and grade performance are independent.

B: the data do not provide sufficient evidence to conclude that extra-curricular participation and grade performance are independent.

C: the data provide sufficient evidence to conclude that extra-curricular participation and grade performance are related.

D: the data do not provide sufficient evidence to conclude that extra-curricular participation and grade performance are related.

Feedback

A : 0

✗ That is not quite right. Notice that the p-value is below .05. This implies that the data provide sufficient evidence to reject the null hypothesis. Consider the remaining options. (C) is the right answer.

B : 0

✗ That is not quite right. Notice that the p-value is below .05. This implies that the data provide sufficient evidence to reject the null hypothesis. Consider the remaining options. (C) is the right answer.

C : 10

✓ Good job! The p-value (0.046) is sufficiently small (less than .05) to reject H_0 in favor of the alternative hypothesis and conclude that extra-curricular activity and grade performance are related.

D : 0

✗ That is not quite right. Notice that the p-value is below .05. This implies that the data provide sufficient evidence to reject the null hypothesis. Consider the remaining options. (C) is the right answer.

Question (7)

Which of the following facts should make you the most worried about the reliability of the results of the test in this case?

A: The 112 students were all students in a chemical engineering course and not a random sample from the entire student body.

B: Two of the six observed counts are less than 5.

C: Not all of the cells' contributions to the chi-squared statistic are greater than 1.

D: Two of the six expected counts are less than 5.

Feedback

A : 0

✗ That is not quite right. Although these students are all from the same course, they are still randomly chosen. Consider the remaining options. (D) is the right answer.

B : 0

✗ That is not quite right. Rather than two of the six observed counts being less than 5, two of the six expected counts being less than 5 should worry you. Consider the remaining options. (D) is the right answer.

C : 0

✗ That is not quite right. This just implies that some cells have expected counts and observed counts that are very close in value. (D) is the right answer.

D : 10

✓ Good job! Since two of the six (or 33% of the) expected counts are less than 5 we should be concerned about the reliability of the results of the test. The fact that the random sample was not taken from among the entire university's students but from a smaller sampling frame of students in a chemical engineering course should not harm the reliability of the test results (since the sample is still random) but should just affect how general our conclusions can be.